

Sequence Appendix 1

Untitled Sequence # 2 -> 1-phase Translation

DNA sequence 1983 b.p. TTCCATCTAACC ... ATGGTGGGCGCC linear

1/1 31/11
 TTC CAT CTA ACC ACA CGT AAC GGA GAA CCA CAC ATG ATC GTC AGT AGA CAA GAG AAA GGG
 F H L T T R N G E P H M I V S R Q E K G
 61/21 91/31
 AAA AGT CTT TTG TTT AAA ACA GAG GAT GGC GTG AAC ATG TGC ACC CTC ATG GCC ATG GAC
 K S L L F K T E D G V N M C T L M A M D
 121/41 151/51
 CTT GGT GAA TTG TGT GAA GAC ACA ATC ACG TAC AAG TGT CCC CTT CTC AGG CAG AAT GAG
 L G E L C E D T I T Y K C P L L R Q N E
 181/61 211/71
 CCA GAA GAC ATA GAC TGC TGG TGC AAC TCC ACG TCC ACG TGG GTA ACC TAT GGG ACT TGT
 P E D I D C W C N S T S T W V T Y G T C
 241/81 271/91
 ACC ACC ACG GGA GAA CAT AGA AGA GAA AAA AGA TCA GTG GCA CTC GTT CCA CAT GTG GGA
 T T T G E H R R E K R S V A L V P H V G
 301/101 331/111
 ATG GGA CTC GAG ACG CGA ACT GAA ACA TGG ATG TCA TCA GAA GGG GCT TGG AAA CAT GCC
 M G L E T R T E T W M S S E G A W K H A
 361/121 391/131
 CAG AGA ATT GAA ATT TGG ATC CTG AGA CAT CCA GGC TTC ACC ATA ATG GCA GCA ATC CTG
 Q R I E I W I L R H P G F T I M A A I L
 421/141 451/151
 GCA TAC ACC ATA GGG ACG ACA CAT TTC CAG AGA GCA CTG ATT TTC ATC TTA CTG ACA GCT
 A Y T I G T T H F Q R A L I F I L L T A
 481/161 511/171
 GTC GCT CCT TCA ATG ACA ATG CGT TGC ATA GGA ATA TCA AAT AGA GAC TTT GTA GAA GGG
 V A P S M T M R C I G I S N R D F V E G
 541/181 571/191
 GTT TCA GGA GGA AGC TGG GTT GAC ATA GTC TTA GAA CAT GGA AGC TGT GTG ACG ACG ATG
 V S G G S W V D I V L E H G S C V T T M
 601/201 631/211
 GCA AAA AAC AAA CCA ACA TTG GAT TTT GAA CTG ATA AAA ACA GAA GCC AAA CAG CCT GCC
 A K N K P T L D F E L I K T B A K Q P A
 661/221 691/231
 ACC CTA AGG AAG TAC TGT ATA GAG GCA AAG CTA ACC AAC ACA ACA ACA GAA TCT CGT TGC
 T L R K Y C I E A K L T N T T T E S R C
 721/241 751/251
 CCA ACA CAA GGG GAA CCC ACG CTA AAT GAA GAG CAG GAT AAA AGG TTC GTC TGC AAA CAC
 P T Q G E P S L N E E Q D K R F V C K H
 781/261 811/271
 TCC ATG GTA GAC AGA GGA TGG GGA AAT GGA TGT GGA TTA TTT GGA AAG GGA GGC ATT GTG
 S M V D R G W G N G C G L F G K G G I V
 841/281 871/291
 ACC TGT GCT ATG TTC ACA TGC AAA AAG AAC ATG GAG GGA AAA GTT GTG CAG CCA GAA AAC
 T C A M F T C K K N M E G K V V Q P E N
 901/301 931/311
 TTG GAA TAC ACC ATT GTG GTA ACA CCC CAC TCA GGG GAA GAG CAT GCG GTC GGA AAT GAC
 L E Y T I V V T P H S G E E H A V G N D
 961/321 991/331
 ACA GGA AAA CAT GGC AAG GAA ATC AAA GTA ACA CCA CAG AGT TCC ATC ACA GAA GCA GAA
 T G K H G K E I K V T P Q S S I T E A E
 1021/341 1051/351
 TTG ACA GGT TAT GGC ACT GTC ACG ATG GAG TGC TCT CCG AGA ACA GGC CTC GAC TTC AAT
 L T G Y G T V T M E C S P R T G L D F N
 1081/361 1111/371
 GAG ATG GTG TTG CTG CAG ATG GAA AAT AAA GCT TGG CTG GTG CAT AGG CAA TGG TTC CTA
 E M V L L Q M E N K A W L V H R Q W F L
 1141/381 1171/391
 GAC CTG CCG TTA CCA TGG CTG CCC GGA GCG GAC ACA CAA GCG TCA AAT TGG ATA CAA AAA
 D L P L P W L P G A D T Q G S N W I Q K
 1201/401 1231/411
 GAA ACA TTG GTC ACT TTC AAA AAT CCT CAT GCG AAG AAA CAG GAT GTT GTT GTT TTA GGA
 E T L V T F K N P H A K K Q D V V V L G
 1261/421 1291/431
 TCC CAA GAA GCG GCG ATG CAC ACA GCA CTC ACA GCG GCG ACA GAA ATC CAA ATG TCA TCA
 S Q E G A M H T A L T G A T E I Q M S S

 JC542 U.S. PTO
 09/12/1587
 07/23/98

Untitled Sequence # 2 -> 1-phase Translation

1321/441
 GGA AAC TTA CTC TTC ACA GGA CAT CTC AAG TGC AGG CTG AGA ATG GAC AAG CTA CAG CTC
 G N L L F T G H L K C R L R M D K L Q L
 1381/461
 AAA GGA ATG TCA TAC TCT ATG TGC ACA GGA AAG TTT AAA GTT GTG AAG GAA ATA GCA GAA
 R G M S Y S M C T G K F K V V K E I A E
 1441/481
 ACA CAA CAT GGA ACA ATA GTT ATC AGG GTG CAG TAT GAA GGG GAC GGC TCT CCA TGT AAA
 T Q H G T I V I R V Q Y E G D G S P C K
 1501/501
 ATC OCT TTT GAG ATA ATG GAT TTG GAA AAA AGA CAT GTC TTA GGT CGC CTG ATC ACA GTC
 I P F E I M D L E K R H V L G R L I T V
 1561/521
 AAC CCA ATT GTG ACA GAA AAA GAT AGC CCA GTC AAC ATA GAA GCA GAA OCT CCA TTC GGA
 N P I V T E K D S P V N I E A E P P P G
 1621/541
 GAC AGC TAC ATC ATC ATA GGA GTA GAG CCG GGA CAA CTG AAG CTC AAC TGG TTT AAG AAA
 D S Y I I I G V E P G Q L K L N W F K K
 1681/561
 GGA AGT TCT ATC GGC CAA ATG TTT GAG ACA ACA ATG AGG GGG GCG AAG AGA ATG GCC ATT
 G S S I G Q M P E T T M R G A K R M A I
 1741/581
 TTG GGT GAC ACA GCC TGG GAT TTT GGA TCC CTG GGA GGA GTG TTT ACA TCT ATA GGA AAA
 L G D T A W D F G S L G G V F T S I G K
 1801/601
 GCC CTC CAC CAA GTC TTT GGA GCA ATC TAT GGA GCT GCC TTC AGT GGG GTC TCA TGG ACT
 A L H Q V F G A I Y G A A F S G V S W T
 1861/621
 ATG AAA ATC CTC ATA GGA GTC ATT ATC ACA TGG ATA GGA ATG AAT TCA CCG AGC ACC TCA
 M K I L I G V I I T W I G M N S R S T S
 1921/641
 CTG TCT GTG TCA CTA GTA TTG GTG GGA GTC GTG ACG CTG TAT TTG GGA GTT ATG GTG GGC
 L S V S L V L V G V V T L Y L G V M V G
 1981/661
 GCC
 A

10 20 30 40
AGTAAATCCTGTGTGCTAATTGAGGTGCATTGGTCTGCAA 40
ATCGAGTTGCTAGGCAATAAACACATTTGGATTAATTTTA 80
ATCGTTCGTTGAGCGATTAGCAGAGAACTGACCAGAACAT 120
GTCTGGTCGTAAAGCTCAGGGAAAAACCCTGGGCGTCAAT 160
ATGGTACGACGAGGAGTTCGCTCCTTGTCAAACAAAATAA 200

210 220 230 240
AACAAAAACAAAACAAATTGGAACAGACCTGGACCTTC 240
AAGAGGTGTTCAAGGATTTATCTTTTTCTTTTGTTC AAC 280
ATTTTGACTGGAAAAAGATCACAGCCCACCTAAAGAGGT 320
TGTGGAAAATGCTGGACCCAAGACAAGGCTTGCTGTTCT 360
AAGGAAAGTCAAGAGAGTGGTGGCCAGTTGATGAGAGGA 400

410 420 430 440
TTGTCCTCAAGGAAACGCCGTTCCCATGATGTTCTGACTG 440
TGCAATTCCTAATTTTGGGAATGCTGTTGATGACGGGTGG 480
AATGAAGTTGTGCAATTTCCAGGGGAAGCTTTTGATGACC 520
ATCAACAACACGGACATTGCAGACGTTATCGTGATTCCCA 560
CCTCAAAGGAGAGAACAGATGTTGGGTTCGGGCAATCGA 600

610 620 630 640
CGTCGGCTACATGTGTGAGGACACTATCACGTACGAATGT 640
CCTAAGCTTACCATGGGCAATGATCCAGAGGATGTGGATT 680
GCTGGTGTGACAACCAAGAAGTCTACGTCCAATATGGACG 720
GTGCACGCGGACCAGGCATTCCAAGCGAAGCAGGAGATCC 760
GTGTCGGTCCAAACACATGGGGAGAGTTCACTAGTGAATA 800

810 820 830 840
AAAAAGAGGCTTGGCTGGATTCAACGAAAGCCACACGATA 840
TCTCATGAAAAGTGAAGTGGATCATAAGGAATCCTGGC 880
TATGCTTTCCTGGCGGCGGTACTTGGCTGGATGCTTGGCA 920
GTAACAACGGTCAACGCGTGGTATTTACCATCCTCCTGCT 960
GTTGGTCGCTCCGGCTTACAGTTTTAATTGTCTGGGAATG 1000

1010 1020 1030 1040
GGCAATCGTGACTTCATAGAAGGAGCCAGTGGGGCCACTT 1040
GGGTGGACTTGGTGCTAGAAGGAGACAGCTGCTTGACAAT 1080
CATGGCAAACGACAAACCAACATTGGACGTCCGCATGATT 1120
AACATCGAAGCTAGCCAACTTGCTGAGGTCAGAAAGTTACT 1160
GCTATCATGCTTCAGTCACTGACATCTCGACGGTGGCTCG 1200

1210 1220 1230 1240

GTGCCCCACGACTGGAGAAGCCCACAACGAGAAGCGAGCT 1240
GATAGTAGCTATGTGTGCAAACAAGGCTTCACTGACCGTG 1280
GGTGGGGCAACGGATGTGGATTTTTCGGGAAGGGAAGCAT 1320
TGACACATGTGCAAAATTCTCCTGCACCAGTAAAGCGATT 1360
GGGAGAACAATCCAGCCAGAAAACATCAAATACAAAGTTG 1400

1410 1420 1430 1440

GCATTTTGTGCATGGAACCACCACTTCGGAAAACCATGG 1440
GAATTATTCAGCGCAAGTTGGGGCGTCCCAGGCGGCAAAG 1480
TTTACAGTAACACCCAATGCTCCTTCGGTAGCCCTCAAAC 1520
TTGGTGACTACGGAGAAGTCACACTGGACTGTGAGCCAAG 1560
GAGTGGACTGAACACTGAAGCGTTTTACGTCATGACCGTG 1600

1610 1620 1630 1640

GGGTCAAAGTCATTTCTGGTCCATAGGGAGTGGTTTCATG 1640
ACCTCGCTCTCCCCTGGACGTCCCCTTCGAGCACAGCGTG 1680
GAGAAACAGAGAACTCCTCATGGAATTTGAAGGGGCGCAC 1720
GCCACAAAACAGTCCGTTGTTGCTCTTGGGTACAGGAAG 1760
GAGGCCTCCATCATGCGTTGGCAGGAGCCATCGTGGTGA 1800

1810 1820 1830 1840

GTACTCAAGCTCAGTGATGTTAACATCAGGCCACCTGAAA 1840
TGTAGGCTGAAAATGGACAAACTGGCTCTGAAAGGCACAA 1880
CCTATGGCATGTGTACAGAAAAATTCTCGTTCGCGAAAAA 1920
TCCGGTGGACACTGGTCACGGAACAGTTGTCATTGAACTC 1960
TCCTACTCTGGGAGTGATGGCCCCCTGCAAAATTCCGATTG 2000

2010 2020 2030 2040

TTTCCGTTGCGAGCCTCAATGACATGACCCCCGTTGGGCG 2040
GCTGGTGACAGTGAACCCCTTCGTCGCGACTTCCAGTGCC 2080
AACTCAAAGGTGCTGGTCGAGATGGAACCCCTTCGGAG 2120
ACTCCTACATCGTAGTTGGAAGGGGAGACAAGCAGATCAA 2160
CCACCATTGGCACAAGCTGGAAGCACGCTGGGCAAGGCC 2200

2210 2220 2230 2240

TTTTCAACAACCTTGAAGGGAGCTCAAAGACTGGCAGCGT 2240
TGGGCGACACAGCCTGGGACTTTGGCTCTATTGGAGGGGT 2280
CTTCAACTCCATAGGAAGAGCCGTTACCAAGTGTGTTGGT 2320
GGTGCCTTCAGAACTCTTTGGGGGAATGTCTTGGATCA 2360
CACAAGGGCTAATGGGTGCCCTACTGCTCTGGATGGGCGT 2400

2410 2420 2430 2440

CAACGCACGAGACCGATCAATTGCTTTGGCCTTCTTAGCC 2440
ACAGGAGGTGTGCTCGTGTTCTTAGCGACCAATGTGGGCG 2480
CCGATCAAGGATGCGCCATCAACTTTGGCAAGAGAGAGCT 2520
CAAGTGCGGAGATGGTATCTTCATATTTAGAGACTCTGAT 2560
GACTGGCTGAACAAGTACTCATACTATCCAGAAGATCCTG 2600

2610 2620 2630 2640

TGAAGCTTGCATCAATAGTGAAAGCCTCTTTTGAAGAAGG 2640
GAAGTGTGGCCTAAATTCAGTTGACTCCCTTGAGCATGAG 2680
ATGTGGAGAAGCAGGGCAGATGAGATCAATGCCATTTTTG 2720
AGGAAAACGAGGTGGACATTTCTGTTGTCGTGCAGGATCC 2760
AAAGAATGTTTACCAGAGAGGAACATCATCCATTTTCCAGA 2800

2810 2820 2830 2840

ATTCGGGATGGTCTGCAGTATGGTTGGAAGACTTGGGGTA 2840
AGAACCTTGTGTTCTCCCCAGGGAGGAAGAATGGAAGCTT 2880
CATCATAGATGGAAAGTCCAGGAAAGAATGCCCGTTTTCA 2920
AACCGGGTCTGGAATTCTTTCCAGATAGAGGAGTTTGGGA 2960
CGGGAGTGTTACACACACGCGTGACATGGACGCAGTCTT 3000

3010 3020 3030 3040

TGAATACACCATAGACTGCGATGGATCTATCTTGGGTGCA 3040
GCGGTGAACGGAAAAAAGAGTGCCCATGGCTCTCCAACAT 3080
TTTGGATGGGAAGTCATGAAGTAAATGGGACATGGATGAT 3120
CCACACCTTGGAGGCATTAGATTACAAGGAGTGTGAGTGG 3160
CCACTGACACATACGATTGGAACATCAGTTGAAGAGAGTG 3200

3210 3220 3230 3240

AAATGTTTCATGCCGAGATCAATCGGAGGCCAGTTAGCTC 3240
TCACAATCATATCCCTGGATACAAGGTTGAGACGAACGGA 3280
CCTTGGATGCAGGTACCACTAGAAGTGAAGAGAGAAGCTT 3320
GCCCAGGGACTAGCGTGATCATTGATGGCAACTGTGATGG 3360
ACGGGGAAAAATCAACCAGATCCACCACGGATAGCGGGAAA 3400

3410 3420 3430 3440

GTTATTCCTGAATGGTGTGCGCTCCTGCACAATGCCGC 3440
CTGTGAGCTTCCATGGTAGTGATGGGTGTTGGTATCCCAT 3480
GGAAATTAGGCCAAGGAAAACGCATGAAAGCCATCTGGTG 3520
CGCTCCTGGGTTACAGCTGGAGAAATACATGCTGTCCCTT 3560
TTGGTTTGGTGAGCATGATGATAGCAATGGAAGTGGTCCT 3600

3610 3620 3630 3640

AAGGAAAAGACAGGGACCAAGCAAATGTTGGTTGGAGGA 3640
GTAGTGCTCTTGGGAGCAATGCTGGTCGGGCAAGTAACTC 3680
TCCTTGATTTGCTGAACTCACAGTGGCTGTGGGATTGCA 3720
TTTCCATGAGATGAACAATGGAGGAGACGCCATGTATATG 3760
GCGTTGATTGCTGCCTTTTCAATCAGACCAGGGCTGCTCA 3800

3810 3820 3830 3840

TCGGCTTTGGGCTCAGGACCCTATGGAGCCCTCGGGAACG 3840
CCTTGTGCTGACCCTAGGAGCAGCCATGGTGAGATTGCC 3880
TTGGGTGGCGTGATGGGCGGCCTGTGGAAGTATCTAAATG 3920
CAGTTTCTCTCTGCATCCTGACAATAAATGCTGTTGCTTC 3960
TAGGAAAGCATCAAATACCATCTTGCCCCTCATGGCTCTG 4000

4010 4020 4030 4040

TTGACACCTGTCACTATGGCTGAGGTGAGACTTGCCGCAA 4040
TGTTCTTTTGTGCCATGGTTATCATAGGGGTCCTTCACCA 4080
GAATTTCAAGGACACCTCCATGCAGAAGACTATACCTCTG 4120
GTGGCCCTCACACTCACATCTTACCTGGGCTTGACACAAC 4160
CTTTTTTGGGCCTGTGTGCATTTCTGGCAACCCGCATATT 4200

4210 4220 4230 4240

TGGGCGAAGGAGTATCCCAGTGAATGAGGCACTCGCAGCA 4240
GCTGGTCTAGTGGGAGTGCTGGCAGGACTGGCTTTTCAGG 4280
AGATGGAGAACTTCCTTGGTCCGATTGCAGTTGGAGGACT 4320
CCTGATGATGCTGGTTAGCGTGGCTGGGAGGGTGGATGGG 4360
CTAGAGCTCAAGAAGCTTGGTGAAGTTTCATGGGAAGAGG 4400

4410 4420 4430 4440

AGGCGGAGATCAGCGGGAGTTCCGCCCCGCTATGATGTGGC 4440
ACTCAGTGAACAAGGGGAGTTCAAGCTGCTTTCTGAAGAG 4480
AAAGTGGCATGGGACCAGGTTGTGATGACCTCGCTGGCCT 4520
TGGTTGGGGCTGCCCTCCATCCATTTGCTCTTCTGCTGGT 4560
CCTTGCTGGGTGGCTGTTTCATGTCAGGGGAGCTAGGAGA 4600

4610 4620 4630 4640

AGTGGGGATGTCTTGTGGGATATTCCCCTCCTAAGATCA 4640
TCGAGGAATGTGAACATCTGGAGGATGGGATTTATGGCAT 4680
ATTCCAGTCAACCTTCTTGGGGGCCTCCCAGCGAGGAGTG 4720
GGAGTGGCACAGGGAGGGGTGTTCCACACAATGTGGCATG 4760
TCACAAGAGGAGCTTTCCTTGTTCAGGAATGGCAAGAAGTT 4800

4810 4820 4830 4840

GATTCCATCTTGGGCTTCAGTAAAGGAAGACCTTGTGCGC 4840
TATGGTGGCTCATGGAAGTTGGAAGGCAGATGGGATGGAG 4880
AGGAAGAGGTCCAGTTGATCGCGGCTGTTCCAGGAAAGAA 4920
CGTGGTCAACGTCCAGACAAAACCGAGCTTGTTCAAAGTG 4960
AGGAATGGGGGAGAAATCGGGGCTGTCGCTCTTGAATATC 5000

5010 5020 5030 5040

CGAGTGGCACTTCAGGATCTCCTATTGTTAACAGGAACGG 5040
AGAGGTGATTGGGCTGTACGGCAATGGCATCCTTGTGCGGT 5080
GACAACTCCTTCGTGTCCGCCATATCCCAGACTGAGGTGA 5120
AGGAAGAAGGAAAGGAGGAGCTCCAAGAGATCCCGACAAT 5160
GCTAAAGAAAGGAATGACAACTGTCCTTGATTTTCATCCT 5200

5210 5220 5230 5240

GGAGCTGGGAAGACAAGACGTTTCTCCACAGATCTTGG 5240
CCGAGTGCGCACGGAGACGCTTGCCTACTCTTGTGTTGGC 5280
CCCCACCAGGGTTGTTCTTTCTGAAATGAAGGAGGCTTTT 5320
CACGGCCTGGACGTGAAATTCACACACAGGCTTTTTCCG 5360
CTCACGGCAGCGGGAGAGAAGTCATTGATGCCATGTGCCA 5400

5410 5420 5430 5440

TGCCACCCTAACTTACAGGATGTTGGAACCAACTAGGGTT 5440
GTTAACTGGGAAGTGATCATTATGGATGAAGCCCATTTTT 5480
TGGATCCAGCTAGCATAGCCGCTAGAGGTTGGGCAGCGCA 5520
CAGAGCTAGGGCAAATGAAAGTGCAACAATCTTGATGACA 5560
GCCACACCGCCTGGGACTAGTGATGAATTTCCACATTCAA 5600

5610 5620 5630 5640

ATGGTGAAATAGAAGATGTTCAAACGGACATACCCAGTGA 5640
GCCCTGGAACACAGGGCATGACTGGATCCTGGCTGACAAA 5680
AGGCCCACGGCATGGTTCCTTCCATCCATCAGAGCTGCAA 5720
ATGTCATGGCTGCCTCTTTGCGTAAGGCTGGAAAGAGTGT 5760
GGTGGTCCTGAACAGGAAAACCTTTGAGAGAGAATACCCC 5800

5810 5820 5830 5840

ACGATAAAGCAGAAGAAACCTGACTTTATATTGGCCACTG 5840
ACATAGCTGAAATGGGAGCCAACCTTTGCGTGGAGCGAGT 5880
GCTGGATTGCAGGACGGCTTTTAAGCCTGTGCTTGTGGAT 5920
GAAGGGAGGAAGGTGGCAATAAAAGGGCCACTTCGTATCT 5960
CCGCATCCTCTGCTGCTCAAAGGAGGGGGCGCATTGGGAG 6000

6010 6020 6030 6040

AAATCCCAACAGAGATGGAGACTCATACTACTATTCTGAG 6040
CCTACAAGTGAAAATAATGCCCACCACGTCTGCTGGTTGG 6080
AGGCCTCAATGCTCTTGGACAACATGGAGGTGAGGGGTGG 6120
AATGGTCGCCCCACTCTATGGCGTTGAAGGAACTAAAACA 6160
CCAGTTTCCCCTGGTGAAATGAGACTGAGGGATGACCAGA 6200

6210 6220 6230 6240

GGAAAGTCTTCAGAGAACTAGTGAGGAATTGTGACCTGCC 6240
CGTTTGGCTTTCGTGGCAAGTGGCCAAGGCTGGTTTGAAG 6280
ACGAATGATCGTAAGTGGTGTGTTTGAAGGCCCTGAGGAAC 6320
ATGAGATCTTGAATGACAGCGGTGAAACAGTGAAGTGCAG 6360
GGCTCCTGGAGGAGCAAAGAAGCCTCTGCGCCCAAGGTGG 6400

6410 6420 6430 6440

TGTGATGAAAGGGTGTCTATCTGACCAGAGTGCGCTGTCTG 6440
AATTTATTAAGTTTGCTGAAGGTAGGAGGGGAGCTGCTGA 6480
AGTGCTAGTTGTGCTGAGTGAACCTCCCTGATTTCCTGGCT 6520
AAAAAAGGTGGAGAGGCAATGGATACCATCAGTGTGTTCC 6560
TCCACTCTGAGGAAGGCTCTAGGGCTTACCGCAATGCACT 6600

6610 6620 6630 6640

ATCAATGATGCCTGAGGCAATGACAATAGTCATGCTGTTT 6640
ATACTGGCTGGACTACTGACATCGGGAATGGTCATCTTTT 6680
TCATGTCTCCCAAAGGCATCAGTAGAATGTCTATGGCGAT 6720
GGGCACAATGGCCGGCTGTGGATATCTCATGTTCTTGGGA 6760
GGCGTCAAACCCACTCACATCTCCTATGTCATGCTCATAT 6800

6810 6820 6830 6840

TCTTTGTCCTGATGGTGGTTGTGATCCCCGAGCCAGGGCA 6840
ACAAAGGTCCATCCAAGACAACCAAGTGGCATAACCTCATT 6880
ATTGGCATCCTGACGCTGGTTTCAGCGGTGGCAGCCAACG 6920
AGCTAGGCATGCTGGAGAAAACCAAGAGGACCTCTTTGG 6960
GAAGAAGAACTTAATTCATCTAGTGCTTCACCCTGGAGT 7000

7010 7020 7030 7040

TGGCCGGATCTTGACCTGAAGCCAGGAGCTGCCTGGACAG 7040
TGTACGTTGGCATTGTTACAATGCTCTCTCCAATGTTGCA 7080
CCACTGGATCAAAGTCGAATATGGCAACCTGTCTCTGTCT 7120
GGAATAGCCCAGTCAGCCTCAGTCCTTTCTTTTCATGGACA 7160
AGGGGATACCATTTCATGAAGATGAATATCTCGGTCATAAT 7200

7210 7220 7230 7240

GCTGCTGGTCAGTGGCTGGAATTCAATAACAGTGATGCCT 7240
CTGCTCTGTGGCATAGGGTGCGCCATGCTCCACTGGTCTC 7280
TCATTTTACCTGGAATCAAAGCGCAGCAGTCAAAGCTTGC 7320
ACAGAGAAGGGTGTTCATGGCGTTGCCAAGAACCCTGTG 7360
GTTGATGGGAATCCAACAGTTGACATTGAGGAAGCTCCTG 7400

7410 7420 7430 7440

AAATGCCTGCCCTTTATGAGAAGAACTGGCTCTATATCT 7440
CCTTCTTGCTCTCAGCCTAGCTTCTGTTGCCATGTGCAGA 7480
ACGCCCTTTTCATTGGCTGAAGGCATTGTCCTAGCATCAG 7520
CTGCCCTTAGGGCCGCTCATAGAGGGAAACACCAGCCTTCT 7560
TTGGAATGGACCCATGGCTGTCTCCATGACAGGAGTCATG 7600

7610 7620 7630 7640

AGGGGGAATCACTATGCTTTTGTGGGAGTCATGTACAATC 7640
TATGGAAGATGAAAACCTGGACGCCGGGGGAGCGCAATGG 7680
AAAAACTTTGGGTGAAGTCTGGAAGAGGGAACCTGAATCTG 7720
TTGGACAAGCGACAGTTTGAGTTGTATAAAAGGACCGACA 7760
TTGTGGAGGTGGATCGTGATACGGCACGCAGGCATTTGGC 7800

7810 7820 7830 7840

CGAAGGGAAGGTGGACACCGGGGTGGCGGTCTCCAGGGGG 7840
ACCGCAAAGTTAAGGTGGTTCCATGAGCGTGGCTATGTCA 7880
AGCTGGAAGGTAGGGTGATTGACCTGGGGTGTGGCCGCGG 7920
AGGCTGGTGTTACTACGCTGCTGCGCAAAGGAAGTGAGT 7960
GGGGTCAAAGGATTTACTCTTGGAAGAGACGGCCATGAGA 8000

8010 8020 8030 8040

AACCCATGAATGTGCAAAGTCTGGGATGGAACATCATCAC 8040
CTTCAAGGACAAAACCTGATATCCACCGCCTAGAACCAGTG 8080
AAATGTGACACCCTTTTGTGTGACATTGGAGAGTCATCAT 8120
CGTCATCGGTCACAGAGGGGGAAAGGACCGTGAGAGTTCT 8160
TGATACTGTAGAAAAATGGCTGGCTTGTGGGGTTGACAAC 8200

8210 8220 8230 8240

TTCTGTGTGAAGGTGTTAGCTCCATACATGCCAGATGTTC 8240
TTGAGAAACTGGAATTGCTCCAAAGGAGGTTTGGCGGAAC 8280
AGTGATCAGGAACCCTCTCTCCAGGAATTCCACTCATGAA 8320
ATGTACTACGTGTCTGGAGCCCCGAGCAATGTCACATTTA 8360
CTGTGAACCAACATCCCGCCTCCTGATGAGGAGAATGAG 8400

8410 8420 8430 8440

GCGTCCAACCTGGAAAAGTGACCCTGGAGGCTGACGTCATC 8440
CTCCCAATTGGGACACGCACTGTTGAGACAGACAAGGGAC 8480
CCCTGGACAAAGAGGCCATAGAAGAAAGGGTTGAGAGGAT 8520
AAAATCTGAGTACATGACCTCTTGGTTTTATGACAATGAC 8560
AACCCCTACAGGACCTGGCACTACTGTGGCTCCTATGTCA 8600

8610 8620 8630 8640

CAAAAACCTCCGGAAGTGCGGCGAGCATGGTAAATGGTGT 8640
TATTAATAATTCTGACATATCCATGGGACAGGATAGAGGAG 8680
GTCACAAGAATGGCAATGACTGACACAACCCCTTTTGGAC 8720
AGCAAAGAGTGTTTAAAGAAAAAGTTGACACCAGAGCAAA 8760
GGATCCACCAGCGGGAAGTAGGAAGATCATGAAAGTTGTC 8800

8810 8820 8830 8840

AACAGGTGGCTGTTCCGCCACCTGGCCAGAGAAAAGAACC 8840
CCAGACTGTGCACAAAGGAAGAATTTATTGCAAAAGTCCG 8880
AAGTCATGCAGCCATTGGAGCTTACCTGGAAGAACAAGAA 8920
CAGTGGAAGACTGCCAATGAGGCTGTCCAAGACCCAAAGT 8960
TCTGGGAAGTGGTGGATGAAGAAAGGAAGCTGCACCAACA 9000

9010 9020 9030 9040

AGGCAGGTGTCTGGACTTGTGTGTACAACATGATGGGGAAA 9040
AGAGAGAAGAAGCTGTGAGAGTTTGGGAAAGCAAAGGGAA 9080
GCCGTGCCATATGGTATATGTGGCTGGGAGCGCGGTATCT 9120
TGAGTTTGAGGCCCTGGGATTCTGAATGAGGACCATTGG 9160
GCTTCCAGGGAAAACCTCAGGAGGAGGAGTGGAAGGCATTG 9200

9210 9220 9230 9240

GCTTACAATACCTAGGATATGTGATCAGAGACCTGGCTGC 9240
AATGGATGGTGGTGGATTCTACGCGGATGACACCGCTGGA 9280
TGGGACACGCGCATCACAGAGGCAGACCTTGATGATGAAC 9320
AGGAGATCTTGAACATGAGCCACATCACAAAAAACT 9360
GGCACAAGCAGTGATGGAAATGACATACAAGAACAAGTG 9400

9410 9420 9430 9440

GTGAAAGTGTTGAGACCAGCCCCAGGAGGGAAAGCCTACA 9440
TGGATGTCATAAGTCGACGAGACCAGAGAGGATCCGGGCA 9480
GGTAGTGACTTATGCTCTGAACACCATCACCAACTTGAAA 9520
GTCCAATTGATCAGAATGGCAGAAGCAGAGATGGTGATAC 9560
ATCACCAACATGTTCAAGATTGTGATGAATCAGTTCTGAC 9600

9610 9620 9630 9640
CAGGCTGGAGGCATGGCTCACTGAGCACGGATGTGACAGA 9640
CTGAAGAGGATGGCGGTGAGTGGAGACGACTGTGTGGTCC 9680
GGCCCATCGATGACAGGTTTCGGCCTGGCCCTGTCCCATCT 9720
CAACGCCATGTCCAAGGTTAGAAAGGACATATCTGAATGG 9760
CAGCCATCAAAAGGGTGGAATGATTGGGAGAATGTGCCCT 9800

9810 9820 9830 9840
TCTGTTCCCACCACTTCCATGAACTACAGCTGAAGGATGG 9840
CAGGAGGATTGTGGTGCCTTGCCGAGAACAGGACGAGCTC 9880
ATTGGGAGAGGAAGGGTGTCTCCAGGAAACGGCTGGATGA 9920
TCAAGGAAACAGCTTGCCTCAGCAAAGCCTATGCCAACAT 9960
GTGGTCACTGATGTATTTTCACAAAAGGGACATGAGGCTA 10000

10010 10020 10030 10040
CTGTCATTGGCTGTTTCCTCAGCTGTTCCCACCTCATGGG 10040
TTCCACAAGGACGCACAACATGGTCGATTCATGGGAAAGG 10080
GGAGTGGATGACCACGGAAGACATGCTTGAGGTGTGGAAC 10120
AGAGTATGGATAACCAACAACCCACACATGCAGGACAAGA 10160
CAATGGTGAAAAAATGGAGAGATGTCCCTTATCTAACCAA 10200

10210 10220 10230 10240
GAGACAAGACAAGCTGTGCGGATCACTGATTGGAATGACC 10240
AATAGGGCCACCTGGGCCTCCACATCCATTTAGTCATCC 10280
ATCGTATCCGAACGCTGATTGGACAGGAGAAATACTGA 10320
CTACCTAACAGTCATGGACAGGTATTCTGTGGATGCTGAC 10360
CTGCAACTGGGTGAGCTTATCTGAAACACCATCTAACAGG 10400

10410 10420 10430 10440
AATAACCGGGATACAAACCACGGGTGGAGAACCGGACTCC 10440
CCACAACCTGAAACCGGGATATAAACCACGGCTGGAGAAC 10480
CGGGCTCCGCACTTAAATGAAACAGAAACCGGGATAAAA 10520
ACTACGGATGGAGAACCGGACTCCACACATTGAGACAGAA 10560
GAAGTTGTCAGCCCAGAACCCACACGAGTTTGGCCACTG 10600

10610 10620 10630 10640
CTAAGCTGTGAGGCAGTGCAGGCTGGGACAGCCGACCTCC 10640
AGGTTGCGAAAAACCTGGTTTCTGGGACCTCCACCCAG 10680
AGTAAAAAGAACGGAGCCTCCGCTACCACCCTCCACGTG 10720
GTGGTAGAAAGACGGGGTCTAGAGGTTAGAGGAGACCCTC 10760
CAGGGAACAAATAGTGGGACCATATTGACGCCAGGGAAAG 10800

10810 10820 10830 10840
ACCGGAGTGGTTCTCTGCTTTTCCTCCAGAGGTCTGTGAG 10840
CACAGTTTGCTCAAGAATAAGCAGACCTTTGGATGACAAA 10880
CACAAAACCACT 10892

...

10 20 30 40
MSGRKAQGKTLGVNMVRRGVRSLSNKIKQKTKQIGNRPGP 40
SRGVQGFIFFFLFNILTGKKITAHKRLWKMLDPRQGLAV 80
LRKVKRVASLMRGLSSRKRRSHDVLTVQFLILGMLLMTG 120
GMKLSNFQGKLLMTINNTDIADVIVIPTSKGENRCWVRAI 160
DVGVMCEDTITYECPKLTMGNDPEDVDCWCDNQEVYVQYG 200

210 220 230 240
RCTRTRHSKRSRRSVSVQTHGESSLVNKKEAWLDSTKATR 240
YLMKTENWIIRNPGYAFLAAVLGWMLGSNNGQRVVFTILL 280
LLVAPAYSFNCLGMGNRDFIEGASGATWVDLVLEGDSCLT 320
IMANDKPTLDVRMINIEASQLAEVRSYCYHASVTDISTVA 360
RCPTTGEAHNEKRADSSYVCKQGFTDRGWGNGCGFFGKGS 400

410 420 430 440
IDTCAKFSCTSKAIGRTIQPENIKYKVGIFVHGTTTSENH 440
GNYSAQVGASQAAKFTVTPNAPSVALKLG DYGEVTLDCP 480
RSGLNTEAFYVMTVGSKSFLVHREWFHDLALPWTSPSSTA 520
WRNRELLMEFEGAHATKQSVVALGSQEGGLHHALAGAIIV 560
EYSSSVMLTSGHLKCRCLKMDKLALKGTTYGMCTEKFSFAK 600

610 620 630 640
NPVDTGHGTVVIELSYSGSDGPCKIPIVSVASLNDMTPVG 640
RLVTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGDKQI 680
NHHWHKAGSTLGKAFSTTLKGAQRLAALGDTAWDFGSIGG 720
VFNSIGRAVHQVFVGGAFRTLFGGMSWITQGLMGALLLWMG 760
VNARDRSIALAFLATGGVLVFLATNVGADQGCAINFGKRE 800

810 820 830 840
LKCGDGIFIFRSDDDLKYSYYPEDPVKLASIVKASFEE 840
GKCGLNSVDSLEHEMWRSRADEINAI FEENEVDISVVVQD 880
PKNVYQRGTHPFSRIRDGLQYGWKTWGKNLVFSPGRKNGS 920
FIIDGKSRKECPFSNRVWNSFQIEEFGTGVFTTRVYMDAV 960
FEYTIIDCDGSILGAAVNGKKS AHGSPTFWMG SHEVNGTWM 1000

1010 1020 1030 1040
IHTLEALDYKECEWPLTHTIGTSVEESEMFMPSIGGPVS 1040
SHNHIPGYKVQTNPGWMQVPLEVKREACPGTSVII DGNCD 1080
GRGKSTRSTTDSGKVIPEWCCRSC TMPPVSFHGSDGCWYP 1120
MEIRPRKTHESHLVRSWVTAGEIHAVPFGLVSMMIAMEVV 1160

1210	1220	1230	1240
------	------	------	------

HFHEMNNGGDAMYMALIAAFSIRPGLLIGFGLRTLWSPRE 1240
RLVLTTLGAAMVEIALGGVMGGLWKYLNVAVSLCILTINAVA 1280
SRKASNTILPLMALLTPVTMAEVRLAAMFFCAMVIIGVLH 1320
QNFKDTSMQKTIPLVALTLTSYLGLTQPFLGLCAFLATRI 1360
FGRRSIPVNEALAAAGLVGVLGGLAFQEMENFLGPIAVGG 1400

1410	1420	1430	1440
------	------	------	------

LLMMLVSVAGRVDGLELKKLGEVSWEEAEISGSSARYDV 1440
ALSEQGEFKLLSEEKVPWDQVMTSLALVGAALHPFALLL 1480
VLAWLFWHVRGARRSGDVLWDIPTPKIIECEHLEDGIYG 1520
IFQSTFLGASQRGVGVAGGGVFHTMWHVTRGAFLVRNGKK 1560
LIPSWASVKEDLVAYGGSWKLEGRWDGEEVQLIAAVPGK 1600

1610	1620	1630	1640
------	------	------	------

NVVNVQTKPSLFKVRNGGEIGAVALDYPSTSGSPIVNRN 1640
GEVIGLYGNGILVGDNSFVSAISQTEVKEEGKEELQEIPT 1680
MLKKGMTTVLDFHPGAGKTRRFLPQILAECARRRLRTLVL 1720
APTRVVLSEMKEAFHGLDVKFHTQAFSAHSGGREVIDAMC 1760
HATLTYRMLEPTRVVNWEVIIMDEAHFLDPASIAARGWAA 1800

1810	1820	1830	1840
------	------	------	------

HRARANESATILMTATPPGTSDEFPHSNGEIEDVQTDIPS 1840
EPWNTGHDWILADKRPTAWFLPSIRAANVMAASLRKAGKS 1880
VVVLNRKTFEREYPTIKQKKPDFILATDIAEMGANLCVER 1920
VLD CRTAFKPVLVDEGRKVAIKGPLRISASSAAQRRGRIG 1960
RNPNRDGDSEYSEPTSENNAHVVCWLEASMLLDNMEVRG 2000

2010	2020	2030	2040
------	------	------	------

GMVAPLYGVEGKTPVSPGEMRLRDDQKVFRELVRNCDL 2040
PVWLSWQVAKAGLKTNDRKWCFEGPEEHEILNDSGETVKC 2080
RAPGGAKKPLRPRWC DERVSSDQSALSEFIKFAEGRRGAA 2120
EVLVVLSELPDFLAKKGGEAMDTISVFLHSEEGSRAYRNA 2160
LSMMPEAMTIVMLFILAGLLTSGMVIFFMSPKGISRMSMA 2200

2210	2220	2230	2240
------	------	------	------

MGT MAGCGYLMFLGGVKPTHISYVMLIFFVLMVVVIPEPG 2240
QQRSIQDNQVAYLIIGILTLVSAVAANELGMLEKTKEDLF 2280
GKKNLIPSSASPWSWPDLDLKPGAATVYVGIVTMLS PML 2320
HHWIKVEYGNLSLSGIAQSASVLSFMDKGIPFMKMNISVI 2360
MLLVSGWNSITVMPLLCGIGCAMLHWSLILPGIKAQQSKL 2400

2410	2420	2430	2440
------	------	------	------

AQRRVFHGVAKNPVVDGNPTVDIEEAPEMPALYEKKLALY 2440
LLLALSLASVAMCRTPFSLAEGIVLASAALGPLIEGNTSL 2480
LWNGPMAVSMTGVMRGNHYAFVGVMYNLWKMKTGRRGSAN 2520
GKTLGEVWKRELNLLDKRQFELYKRTDIVEVDRDTARRHL 2560
AEGKVDGTGVAVSRTAKLRWFHERGYVKLEGRVIDLGCGR 2600

2610 2620 2630 2640

GGWCYAAAAQKEVSGVKGFTLGRDGHEKPMNVQSLGWNII 2640
TFKDKTDIHRLEPVKCDTLLCDIGESSSSSVTEGERTVRV 2680
LDTVEKWLACGVDFCVKVLAPYMPDVLEKLELLQRRFGG 2720
TVIRNPLSRNSTHEMYVSGARSNVTFTVNQTSRLLMRRM 2760
RRPTGKVTLLEADVILPIGTRSVETDKGPLDKEAIEERVER 2800

2810 2820 2830 2840

IKSEYMTSWFYDNDNPYRTWHYCGSYVTKTSGSAASMVNG 2840
VIKILTYPWDRIEEVTRMAMTDTTPFGQQRVFKEKVDTRA 2880
KDPPAGTRKIMKVVRWLFRHLAREKNPRLCTKEEFIAGV 2920
RSHAAIGAYLEEQEOWKTANEAVQDPKFWELVDEERKLHQ 2960
QGRCRTCVYNMMGKREKKLSEFGKAKGSRAIYMWLGARY 3000

3010 3020 3030 3040

LEFEALGFLNEDHWASRENSGGGVEGIGLQYLGYVIRDLA 3040
AMDGGGFYADDTAGWDTRITEADLDDEQEILNYMSPHHKK 3080
LAQAVMEMTYKNKVVRPAPGGKAYMDVISRRDQRGSG 3120
QVVTYALNTITNLKVQLIRMAEAEMVIHHQHVQDCDESVL 3160
TRLEAWLTEHGCDRLKRMASGDDCVVRPIDDRFGLALSH 3200

3210 3220 3230 3240

LNAMSKVRKDISEWQPSKGWNDWENVPFCSHHFHELQLKD 3240
GRRIVVPCREQDELIGRGRVSPGNGWMIKETACLSKAYAN 3280
MWSLMYFHKRDMRLLSLAVSSAVPTSWVPQGRTTWSIHGK 3320
GEWMTTEDMLEVWNRVWITNPNHMQDKTMVKKWRDVPYLT 3360
KRQDKLCGSLIGMTNRATWASHIHLVIHRITLIGQEKYT 3400

3410 3420 3430 3440

DYLTVMMDRYSVDADLQLGELI. 3422